Overlapping Clustering of Contextual Bandits with NMF techniques

Le Song **Jin Kyoung Kwon** **Shuang Li**

Abstract

We introduce a novel approach to recommendation based on item clustering using non-negative matrix factorization (NMF) techniques. We propose a new algorithm, OCB (Overlapping Clustering Bandits), that groups items into latent clusters using online user feedbacks and uses learned clusters to make recommendations. By making recommendation at cluster-level instead of at item-level, the algorithm can overcome scalability issues associated with a large number of items without compensating for long-term reward maximization. Also, by performing online clustering of items, the algorithm can learn latent topics associated with items based on user feedbacks.

1 Introduction

With an immense growth of online content generation and consumption, recommendation techniques have received much attention in recent years. Traditional methods such as collaborative filtering, content-based filtering and learning-to-rank methods aim to predict a fixed set of recommendations given training data. However, they fall short at dealing with a more realistic setting where a pool of items and users change over time.

Contextual Multi-Armed Bandits (CMAB) algorithms have been successful at addressing this setting [1] [2] [6]. In CMAB, a learning agent faces a sequential decision making problem of choosing the best item for maximizing rewards, given context (side information). In recommendation systems, context can encode features of users and/or items, enabling systems to deal with dynamically changing users and items. The agent solves the task by learning the relationships between context and observed online bandit feedbacks.

1.1 Our objectives

In this paper, we extend this CMAB setting to (1) solve the scalability problem w.r.t the number of items and to (2) show applicability of bandits in learning latent topics of items based on user feedbacks. We approach (1) by incorporating clustering routine to our proposed algorithm. Clustering is a traditional unsupervised learning technique for grouping n data points into k groups. There are two categories of clustering - *hard* clustering where an object belongs to exactly one cluster and *soft* clustering where an object can belong to multiple clusters to varying degrees. When the clusters overlap, the clusters can further be classified as *overlapping*. We employ a routine for performing clustering with non-negative matrix factorization (NMF) techniques that yield soft, overlapping clustering indicators of items as well as parameters for each cluster. Then, we propose a strategy for recommending items using learned cluster parameters, greatly reducing the decision space of the learning agent that is proportional to the number of items in traditional CMAB methods [8].

To approach (2), we introduce a realistic scenario where items can be tagged with many topics. For example, in a news recommendation system, an article can have multiple tags such as business and technology. The degree of association with each tag may vary; for example, the article can be 80% about business and 20% about technology. Recommendation systems may benefit from

having such fine-grained information to better label their available items. This information, however, is time-consuming to and costly obtain. Our algorithm provides a framework for probabilistically interpreting the learned soft, overlapping clusters of items as latent topics. By having assigned probability to each item-topic association, recommendation systems can benefit from having a rich labeling information about items purely based on online user feedbacks.

2 Related Work

Clustering Bandits In [4] and [11], authors investigate clustering of users based on bandit feedbacks. In an extension to this work [9], authors generalize clustering of users to co-clustering between users and items. These works are based on graph partitioning approaches where graph structure is determined by upper confidence bounds on reward estimation. However, due to unweighted edges and hard partitioning of the graph, learned clusters are interpreted as *hard* clusters without probabilistic interpretation. Our results yield soft, overlapping clusters with probabilistic interpretation. We also exclusively focus on clustering of items and do not necessarily require access to user features in order to perform clustering of users or co-clustering between users and items.

Factorization Bandits In [14], authors investigate online factorization between user and items to develop an item selection strategy. Their factorization technique is motivated by collaborative filtering, while our technique is motivated by NMF algorithms. Moreover, their regret guarantee is dependent on access to a pretrained user relational graph that makes it difficult for the recommendation system deal with new users. Perhaps, more similar to our work in factorization bandits is [12]. The authors perform factorization of a reward matrix into two low-rank matrices using NMF, and propose an ϵ-greedy recommendation strategy that uses the learned low-dimensional structure to balance between exploration and exploitation. We, however, propose a recommendation strategy motivated by upper confidence bound (UCB) family of algorithms that provide stronger theoretical guarantees on the expected regret [7]. We also extend the usage of NMF techniques to learn latent topics of items.

Our work merges ideas from both clustering and factorization bandits. We aim to improve upon separate works by introducing an algorithm that can simultaneously learn soft, overlapping clusters of items and maximize long term rewards.

3 Learning Model

3.1 Problem Formulation

Suppose we have a total number of n items, each with context x_i, $i = 1, \ldots, n$ and $x_i \in \mathbb{R}^d$. We want to develop an algorithm that can: (1) do the online recommendation with the goal to maximize the cumulative reward in the long run; (2) adaptively group the n items into k potentially overlapping clustering to get a better understanding of the latent structure of these items.

Define a small number k of clusters with $k < n$, such that items within the same cluster will yield similar rewards. We assume that the items within the same cluster will share the same unknown parameter vector w_j, $j = 1, \ldots, k$. The actual partition of these items will be unknown to the learner, and will be inferred on the fly.

Consider at the decision epoch t, we get access to all the previous chosen items $i[t]$, the corresponding contexts $x_{i[t]}$, and the corresponding rewards denoted as $y[t]$. To group the items into overlapping clusters, we formulate the problem as

$$\min_{W,H} \| \left(Y - H^\top W^\top X\right) \|_F \tag{1}$$

where $Y \in \mathbb{R}^{n \times t}$ is a reward matrix, $H \in \mathbb{R}^{k \times n}$ is a sparse matrix representing clustering indicators, and $W = \{w_1, \ldots, w_k\}$ are parameters for each of k cluster. The decomposition is visualized in Figure 1.

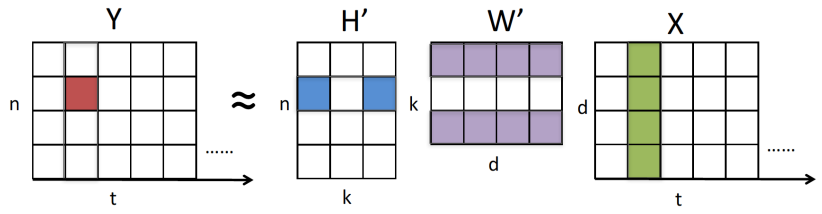

Figure 1

3.2 Applying NMF Techniques for Clustering

The decomposition of Y can be directly solved with NMF techniques. We can reformulate eq. (1) (without Z) as below.

$$\begin{aligned}
\min_{W,H} \|Y - H^\top W^\top X\|_F &= \min_{W,H} \|Y - H^\top (W^\top X)\|_F \\
&= \min_{W,H} \|Y^\top - (W^\top X)^\top (H^\top)^\top\|_F \\
&= \min_{W,H} \|Y^\top - (X^\top W) H\|_F \\
\text{Let } C = X^\top W, C \in \mathbb{R}^{t \times k} & \\
&= \min_{C,H} \|Y^\top - CH\|_F \qquad (2)
\end{aligned}$$

Thus, we can directly estimate Y using NMF techniques. Standard NMF techniques enforce non-negativity constraint on Y, C and H (e.g. they solve $Y_+^T \approx C_+ H_+$). In our case, Y represents reward of n items over t rounds and C represents transformed cluster parameter matrix W. Since rewards and cluster parameters can be negative, we relax non-negativity constraints on Y and C. We do, however, need to enforce the constraint on H whose elements represent cluster memberships of n items in k clusters (and memberships cannot be negative). Two variants of NMF, *Semi-NMF* and *Convex-NMF*, provides techniques for solving $Y_\pm^T \approx C_\pm H_+$ [3]. In our work, we use Semi-NMF algorithm recently introduced in [5] for their relative simplicity in update rules.

3.3 Post-processing NMF results

After we apply Semi-NMF to solve for C and H, we must post-process C and H (1) using W, cluster parameter matrix, to make cluster-level recommendations and (2) using H, cluster indicator matrix, for assigning probabilities to each item-topic association.

3.3.1 Post-processing C

We post-process C and retrieve W, with the following transformations:

$$\begin{aligned}
C &= X^T W \\
XC &= XX^T W \\
(XX^T)^{-1} XC &= (XX^T)^{-1} (XX^T) W \\
(XX^T)^{-1} XC &= W, W \in \mathbb{R}^{d \times m}
\end{aligned}$$

Hence, we can use k columns of W to as cluster parameters to make recommendation.

3.3.2 Post-processing H

Post-processing of H is necessary to retrieve probabilistic interpretation. Namely, we would like $H_{k,i}$ to represent the posterior probability that item i belongs to the k^{th} cluster. An ad-hoc way is to

enforce sum-to-one constraint on columns of H, e.g. $\Sigma_{k=0}^{K} H_{k,i} = 1$ for all item i. However, this is not very rigorous, since NMF solutions are not unique. Suppose $(\hat{C}, \hat{H})$ is a solution of eq. (2). There exists many matrices (A, B) such that $AB^T = I, CA \geq 0, HB \geq 0$. Thus, (CA, CB) is also a solution with the same residue, $||Y^T - CH||$.

To solve the non-uniqueness problem, the following approaches are suggested in [10]. Let z_k be a latent cluster variable. Then, cluster probability for x_i is,

$p(z_k|x_i) \propto (H^T D_C)_{ik}$, D_C =diag($||c_1||, \cdots, ||c_k||$) and $C = (c_1, \cdots, c_k)$.

Therefore, we use the columns of C to rescale the entries in H and enforce sum-to-one constraint to retrieve probabilistic interpretation.

3.4 Making Recommendation

According to our model assumption, the expected reward y_i for item i with observed context x_i is,

$$y_i = \sum_j h^*_{ji} w^{*\top}_j x_i + \epsilon_i \tag{3}$$

where we assume $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ and iid.

We develop a recommendation strategy, motivated by UCB family of algorithms, to trade off exploitation and exploration. In our model, we make recommendation at item level at each timestep t, upon seeing new context $x_{i,t}$, according to

$$a_t \overset{\text{def}}{=} argmax \left(\sum_j \hat{h}_{ji} \hat{w}^{\top}_j x_i + \alpha \cdot VAR \right) \tag{4}$$

where α represents a exploration-exploitation tradeoff parameter and can be chosen by setting $\alpha = 1 + \sqrt{ln(2/\gamma)/2}$ with γ being confidence level [13].

4 Algorithm

Algorithm 1: Overlapping Clustering Bandits (OCB)

Parameters: α, k
Initialize: $Y \in \mathbb{R}^{n \times t}$ with small random numbers
for *t from* 0 *to* T **do**
- Observe context x_i for every item i
- Perform $(\hat{C}, \hat{H}) \leftarrow Y^T$ using `semiNMF` algorithm in [5].
- Retrieve cluster parameter matrix, $W \leftarrow (XX^T)^{-1} X\hat{C}$
- Normalize $\hat{H}$ with l_1 norm
- **for** *every item* i **do**
  - $ucb_i \leftarrow \Sigma_k H[k,i] W[:,k]^T x_i + \alpha \cdot VAR$
- Choose item i^* = argmax ucb_i with ties broken arbitrarily
- Observe a real valued payoff r_t on recommended item a
- $Y[a,t] \leftarrow r_t$

Our algorithm, OCB operates on a round-by-round basis where context $x_t \in \mathbb{R}^d$ is observed every round for every item. Upon observing contexts, the algorithm calls `semiNMF` to factorize reward matrix Y into two latent matrices, C and H. The algorithm then performs post-processing of C and H discussed in Section 3.3 to obtain W and H that can be used for making recommendations. More specifically, the algorithm makes recommendation by estimating upper confidence bound of every item using cluster parameter in a column of W multiplied with item-cluster probabilities in entries of H.

5 Current Work

We are currently working on estimating predictive variance of reward estimation, VAR, that is used in ucb score calculation for every item i in our algorithm (see line 6).

An ad-hoc way is to use the closed-form formula in [8]. Namely, let $A_i = D_i^\top D_i + I_d$ for all arm i, where D_i is a design matrix of dimension $m \times d$ whose rows contain $d-$dimensional contexts observed in previous m rounds. Then, VAR of seeing new context, $x_{t,i}$, can be calculated as $\sqrt{x_{t,i}^\top A_i^{-1} x_{t,i}}$.

In this work, we recognize that a more rigorous approach would involve estimation of propagated error during the updating of $\hat{C}, \hat{H}$ in `semiNMF` algorithm. The difficulty lie in the fact that update rules are nonlinear w.r.t. Y, the matrix that is factorized and contains noise. Currently, we are investigating ways to employ a more robust variance estimation based on first-order Taylor's approximation techniques also known as Delta's method.

6 Conclusion

In this paper, we introduce a novel approach to recommendation based on item soft, overlapping item clustering using NMF techniques. We aim to simultaneously maximize long-term rewards and learn latent topics associated with items based on online user feedbacks. We demonstrate the usage of NMF techniques and appropriate post-processing procedures for obtaining probabilistic interpretation of item-cluster relationships. Moreover, we propose an algorithm that uses learned latent structure to make recommendations.

References

[1] Alekh Agarwal, Daniel J. Hsu, Satyen Kale, John Langford, Lihong Li, and Robert E. Schapire. Taming the monster: A fast and simple algorithm for contextual bandits. *CoRR*, abs/1402.0555, 2014.

[2] Djallel Bouneffouf, Amel Bouzeghoub, and Alda Lopes Gançarski. *Contextual Bandits for Context-Based Information Retrieval*, pages 35–42. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[3] Chris H. Q. Ding, Tao Li, and Michael I. Jordan. Convex and semi-nonnegative matrix factorizations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(1):45–55, January 2010.

[4] Claudio Gentile, Shuai Li, and Giovanni Zappella. Online clustering of bandits. *CoRR*, abs/1401.8257, 2014.

[5] N. Gillis and A. Kumar. Exact and Heuristic Algorithms for Semi-Nonnegative Matrix Factorization. *ArXiv e-prints*, 2014.

[6] Katja Hofmann, Shimon Whiteson, and Maarten de Rijke. Contextual bandits for information retrieval. In *NIPS 2011: Proceedings of the Conference on Neural Information Processing Systems, Workshop on Bayesian Optimization, Experimental Design and Bandits: Theory and Applications*, December 2011.

[7] Volodymyr Kuleshov and Doina Precup. Algorithms for multi-armed bandit problems. *CoRR*, abs/1402.6028, 2014.

[8] Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 661–670, New York, NY, USA, 2010. ACM.

[9] Shuai Li, Claudio Gentile, Alexandros Karatzoglou, and Giovanni Zappella. Data-dependent clustering in exploration-exploitation algorithms. *CoRR*, abs/1502.03473, 2015.

[10] Tao Li and Chris Ding. The relationships among various nonnegative matrix factorization methods for clustering. In *Proceedings of the Sixth International Conference on Data Mining*, ICDM '06, pages 362–371, Washington, DC, USA, 2006. IEEE Computer Society.

[11] Trong T Nguyen and Hady W Lauw. Dynamic clustering of contextual multi-armed bandits. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1959–1962. ACM, 2014.

[12] Rajat Sen, Karthikeyan Shanmugam, Murat Kocaoglu, Alexandros G. Dimakis, and Sanjay Shakkottai. Latent contextual bandits: A non-negative matrix factorization approach. *CoRR*, abs/1606.00119, 2016.

[13] Thomas J. Walsh, Istvan Szita, Carlos Diuk, and Michael L. Littman. Exploring compact reinforcement-learning representations with linear regression. *CoRR*, abs/1205.2606, 2012.

[14] Huazheng Wang, Qingyun Wu, and Hongning Wang. Factorization bandits for interactive recommendation. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 2695–2702, 2017.